

DECOMPOSITION OF THE KANTOROVICH PROBLEM AND WASSERSTEIN DISTANCES ON SIMPLEXES

DANIIL ZAEV

ABSTRACT. Let X be a Polish space, $\mathcal{P}(X)$ be the set of Borel probability measures on X , and $T: X \rightarrow X$ be a homeomorphism. We prove that for the simplex $\text{Dom} \subseteq \mathcal{P}(X)$ of all T -invariant measures, the Kantorovich metric on Dom can be reconstructed from its values on the set of extreme points. This fact is closely related to the following result: the invariant optimal transportation plan is a mixture of invariant optimal transportation plans between extreme points of the simplex. The latter result can be generalized to the case of the Kantorovich problem with additional linear constraints and the class of ergodic decomposable simplexes.

CONTENTS

1. Introduction	1
2. Preliminaries: simplexes and disintegration of measures	5
3. Kantorovich problem with additional linear restrictions	10
4. Measurable selection of optimal transport plans	15
5. Decomposition of Kantorovich problem	17
6. Decomposition of Wasserstein-like distances	20
Acknowledgment	21
References	22

1. INTRODUCTION

In this paper we study an important modification of the Kantorovich problem: the problem of mass transportation with additional linear constraints. The recent developments about general theory of optimal transport are described in [2], [6], [23], [26].

Transport problem with linear constraints appears naturally in several applications, in particular, in financial mathematics (martingale problem) and in infinite-dimensional analysis (invariant transport problem w.r.t. action of some group), see [4], [16], [18], [27].

Many fundamental problems of mass transport theory (e.g., the problem of existence of a transport map) for measures on infinite-dimensional spaces lead us to study transportation of ergodic measures and the relations between the structure of optimal transport plans and ergodic decompositions (see [16] for details).

Date: November 5, 2015.

Key words and phrases. Kantorovich problem, Wasserstein distance, ergodic decomposition, Choquet simplex, invariant measure, disintegration of measures, sufficient statistic, Markov kernel.

In this paper we provide an answer for the question: when is it possible to ergodically decompose an optimal transport plan into optimal plans between decompositions of marginals?

Let us start with an example. Let \mathbb{G} be a nice enough group (locally-compact, amenable), acting continuously on a compact metric space (X, d) . Its action on $X \times X$ is defined in the diagonal way: $g(x_1, x_2) := (g(x_1), g(x_2))$, where g is the action of an element $g \in \mathbb{G}$ on X . Measure μ on (X, d) is called **invariant** w.r.t. \mathbb{G} iff $\mu \circ g^{-1} = \mu$ for every $g \in \mathbb{G}$. Denote by $\mathcal{P}_{\mathbb{G}}(X) \subseteq \mathcal{P}(X)$ the set of all invariant Borel probability measures on (X, d) , equipped with the topology of weak convergence and the corresponding Borel σ -algebra. It is well-known, that $\mathcal{P}_{\mathbb{G}}(X)$ is a compact convex non-empty set. As any other convex set, it contains a subset of all extreme points, i.e. such points, that are midpoints of no line segment with endpoints in $\mathcal{P}_{\mathbb{G}}(X)$. Let us denote this set via $\partial_e(\mathcal{P}_{\mathbb{G}}(X))$ and call it the boundary of $\mathcal{P}_{\mathbb{G}}(X)$. It is clear, that extreme points of $\mathcal{P}_{\mathbb{G}}(X)$ are exactly \mathbb{G} -ergodic measures. It is also known (see [21] for a proof), that $\mathcal{P}_{\mathbb{G}}(X)$ is a simplex. It means that each point of $\mathcal{P}_{\mathbb{G}}(X)$ can be represented in the unique way as a barycenter of a probability Borel measure concentrated on the boundary $\partial_e(\mathcal{P}_{\mathbb{G}}(X))$.

Assume that μ, ν are some given measures from $\mathcal{P}_{\mathbb{G}}(X)$. Consider the following optimization problem, which is usually called Kantorovich problem ([2], [6], [23], [26]):

$$\inf \left\{ \int cd\pi : \pi \in \mathcal{P}(X \times X), \Pr_1(\pi) = \mu, \Pr_2(\pi) = \nu \right\}.$$

Here $c : X \times X \rightarrow \mathbb{R}$ is some continuous bounded below function, $\Pr_k(\pi)$ is the k -th marginal of measure π . One can think about this problem as about a problem of finding an optimal way to transfer one mass distribution (measure μ) to the other one (measure ν) with respect to the transfer cost c . Measures π are called transport plans, and can be thought of as generalized maps from one measure space to the other.

It is quite natural to consider in this situation not the set of all transport plans, but the set of invariant ones w.r.t. the action of \mathbb{G} on $X \times Y$. Then the corresponding the Kantorovich problem has the form:

$$\inf \left\{ \int cd\pi : \pi \in \mathcal{P}_{\mathbb{G}}(X \times X), \Pr_1(\pi) = \mu, \Pr_2(\pi) = \nu \right\}$$

This is an example of Kantorovich problem with additional linear restrictions. This modification of Kantorovich problem was independently formulated and studied in [4] and [27]. In our case an additional restriction is the restriction of invariance.

Since μ and ν are elements of simplex $\mathcal{P}_{\mathbb{G}}(X)$, they can be uniquely represented as follows

$$\mu = \int_{\partial_e(\mathcal{P}_{\mathbb{G}}(X))} \xi_{\alpha} d\tilde{\mu}(\alpha) =: \text{bar}(\tilde{\mu}), \quad \nu = \int_{\partial_e(\mathcal{P}_{\mathbb{G}}(X))} \xi_{\beta} d\tilde{\nu}(\beta) =: \text{bar}(\tilde{\nu}),$$

i.e. as barycenters of some probability measures concentrated on the boundary of $\mathcal{P}_{\mathbb{G}}(X)$: $\tilde{\mu}, \tilde{\nu} \in \mathcal{P}(\partial_e(\mathcal{P}_{\mathbb{G}}(X)))$. It is natural to ask, whether it is possible to decompose Kantorovich

problem in the following way:

$$(1.1) \quad \inf \left\{ \int cd\pi : \pi \in \mathcal{P}_{\mathbb{G}}(X \times X), \Pr_1(\pi) = \mu, \Pr_2(\pi) = \nu \right\} = \\ = \inf \left\{ \int \inf \left\{ \int cd\pi : \pi \in \mathcal{P}_{\mathbb{G}}(X \times X), \Pr_1(\pi) = \xi_\alpha, \Pr_2(\pi) = \xi_\beta \right\} d\tilde{\pi} : \tilde{\pi} \in \tilde{\Pi}(\tilde{\mu}, \tilde{\nu}) \right\},$$

where $\tilde{\Pi}(\tilde{\mu}, \tilde{\nu}) := \{\pi \in \mathcal{P}(\partial_e(\mathcal{P}_{\mathbb{G}}(X)) \times \partial_e(\mathcal{P}_{\mathbb{G}}(X))) : \Pr_1(\pi) = \tilde{\mu}, \Pr_2(\pi) = \tilde{\nu}\}$. In other words, is it possible to split Kantorovich problem into two sub-problems: the first one is to calculate the minimal cost of transportation between any pair of measures **from** the boundary of a simplex, and the second one is to find an optimal way to transport measures **on** the boundary, i.e. measures on the set of ergodic measures? In this paper we describe sufficient conditions for existence of such a decomposition. It appears, that this decomposition result holds for a special kind of simplexes $\text{Dom} \subseteq \mathcal{P}(X)$, known as ergodic decomposable simplexes ([11], [17]). The simplex of invariant measures, described above, is an example of an ergodic decomposable simplex. In particular, the following statement is valid.

Theorem 1.1. Let (X, \mathcal{A}) , (Y, \mathcal{B}) be two polish spaces with Borel σ -algebras with given continuous actions of the group $(\mathbb{Z}, +)$, $\mathcal{A}^{inv} \subseteq \mathcal{A}$, $\mathcal{B}^{inv} \subseteq \mathcal{B}$ be corresponding σ -subalgebras of invariant sets, $c: X \times Y \rightarrow \mathbb{R}$ be some lower semicontinuous and bounded below function. Then the equality (1.1) is valid, where $\tilde{\mu}$ is the restriction of μ on \mathcal{A}^{inv} , and $\tilde{\nu}$ is the restriction of ν on \mathcal{B}^{inv} .

This theorem is a particular case of the main theorem of the paper (Theorem 5.5). For the proof let us consider ergodic decomposition of an optimal invariant transport plan. It is not hard to check, that this decomposition consists of ergodic transport plans with ergodic marginals. It is required to show that almost all plans in this decompositions are optimal. If it is not true, one can replace non-optimal components with the optimal ones and obtain a contradiction. Meanwhile, there is a technical difficulty here: making such a replacement one need to be careful with measurability. The family of transport plans consisting the decomposition should remain to be measurably dependent on marginals. This issue is discussed in details in Section 4 of the paper.

Let $\text{Dom} \subseteq \mathcal{P}(X)$ be a simplex that admits decomposition of a Kantorovich problem. Denote for shortness $E := \partial_e(\text{Dom}) \subset \text{Dom}$. Assume that a distance function $\bar{d}: E \times E \rightarrow \mathbb{R}_{\geq 0}$ is defined on E , and that \bar{d} metrizes the topology of E . If we fix some $p \in [1, \infty)$, we can extend \bar{d} to a distance function \hat{d}_p on Dom via the formula:

$$(1.2) \quad \hat{d}_p(\mu, \nu) := \inf \left\{ \left(\int \bar{d}^p(e_1, e_2) d\pi \right)^{\frac{1}{p}} : \pi \in \mathcal{P}(E \times E), \Pr_1(\pi) = \tilde{\mu}, \Pr_2(\pi) = \tilde{\nu} \right\},$$

where $\mathcal{P}(E \times E)$ is the set of all Borel probability measures on $E \times E$, \Pr_k is an operator sending a measure from $\mathcal{P}(E \times E)$ to its k -th marginal, $\text{bar}(\tilde{\mu}) = \mu$, $\text{bar}(\tilde{\nu}) = \nu$.

Let us recall the definition of the L^p -Wasserstein distance in the case the measures are defined on a compact metric space.

Definition 1.2. Let (X, d) be a compact metric space, $\mathcal{P}(X)$ be a set of all Borel (= Radon) probability measures on it. Then for any $p \in [1, \infty)$ L^p -Wasserstein distance $W_p: \mathcal{P}(X) \times$

$\mathcal{P}(X) \rightarrow \mathbb{R}_{\geq 0}$ is a distance function defined by the formula:

$$(1.3) \quad W_p(\mu, \nu) := \inf \left\{ \left(\int d^p(x, y) d\pi \right)^{\frac{1}{p}} : \pi \in \mathcal{P}(X \times X), \Pr_1(\pi) = \mu, \Pr_2(\pi) = \nu \right\}.$$

Considering invariant Kantorovich problem, it natural to introduce the following analogue L^p -Wasserstein distance on $\text{Dom} := \mathcal{P}_{\mathbb{G}}(X)$:

$$(1.4) \quad W_p^{\mathbb{G}}(\mu, \nu) := \inf \left\{ \left(\int d^p(x, y) d\pi \right)^{\frac{1}{p}} : \pi \in \mathcal{P}_{\mathbb{G}}(X \times X), \Pr_1(\pi) = \mu, \Pr_2(\pi) = \nu \right\}.$$

It is known (but it is not a trivial fact), that in the case d is invariant w.r.t. \mathbb{G} , function $W_p^{\mathbb{G}}$ coincides with W_p on Dom (see Moameni, [18]). We do not assume invariance of a distance d , thus $W_p^{\mathbb{G}} \geq W_p$ on their common domain of definition. It can be checked, that $W_p^{\mathbb{G}}$ actually satisfies every axiom of a distance function on $\mathcal{P}_{\mathbb{G}}(X)$ (see Example 3.14 for the proof).

Question 1.3. Let for some $p \in [1, +\infty)$ the distance $W_p^{\mathbb{G}}$ is defined on Dom . Is it possible to define a distance \bar{d} on $E = \partial_e \text{Dom}$ in such a way, that its extension \hat{d}_p on Dom would coincide with $W_p^{\mathbb{G}}$?

Let $\bar{d}_p^{\mathbb{G}}$ be the restriction of the invariant L^p -Wasserstein distance (defined via formula (1.4)) to $\text{Erg}(X) := \partial_e(\text{Dom})$, i.e.

$$(1.5) \quad (\bar{d}_p^{\mathbb{G}})^p(\xi_{\alpha}, \xi_{\beta}) := \inf \left\{ \int d^p(\xi_{\alpha}, \xi_{\beta}) d\pi : \pi \in \mathcal{P}_{\mathbb{G}}(X \times X), \Pr_1(\pi) = \xi_{\alpha}, \Pr_2(\pi) = \xi_{\beta} \right\}.$$

Then it is true, that

$$W_p^{\mathbb{G}}(\mu, \nu) = \hat{d}_p^{\mathbb{G}}(\mu, \nu)$$

for every $\mu, \nu \in \text{Dom}$. Here $\hat{d}_p^{\mathbb{G}}$ is the extension of $\bar{d}_p^{\mathbb{G}}$ (1.5) from $E := \partial_e(\text{Dom})$ on Dom defined by the formula (1.2).

This fact is a particular case of Theorem 6.1. It is clear, that it is closely related to the decomposition of the Kantorovich problem. Denote via $(\mathcal{P}_{\mathbb{G}}(X), W_p^{\mathbb{G}}(d))$ the space of all invariant probability measures on a metric compact space (X, d) , and equip it with the invariant L^p -Wasserstein distance. Since

$$\partial_e(\text{Dom}) \subseteq \text{Dom} = \mathcal{P}_{\mathbb{G}}(X),$$

one can restrict $W_p^{\mathbb{G}}(d)$ on the subsets Dom and $\partial_e(\text{Dom})$ and obtain metric spaces $(\text{Dom}, W_p^{\mathbb{G}}(d))$ and $(\partial_e(\text{Dom}), W_p^{\mathbb{G}}(d))$ respectively. By the definition of simplex,

$$\text{Dom} \simeq \mathcal{P}(\partial_e(\text{Dom})),$$

where “ \simeq ” is for homeomorphism. Hence one can construct L^p -Wasserstein distance on $\mathcal{P}(\partial_e(\text{Dom}))$ considering $(\partial_e(\text{Dom}), W_p^{\mathbb{G}}(d))$ as the original metric structure. Denote the constructed metric space via $(\mathcal{P}(\partial_e(\text{Dom})), W_p(W_p^{\mathbb{G}}(d)))$. Due to the decomposition result,

$$(\mathcal{P}(\partial_e(\text{Dom})), W_p(W_p^{\mathbb{G}}(d))) = (\text{Dom}, W_p^{\mathbb{G}}(d)),$$

where “ $=$ ” is for isometric isomorphism. We shall call results of this type as decompositions of a Wasserstein distance.

We are going to prove a general result about decompositions of transport plans, such that the described decompositions of Wasserstein distances appear to be its particular cases. The main theorems of this paper are Theorem 5.5 and Theorem 6.1. The first one establishes a decomposition result for a Kantorovich problem, while the second one describes decomposition for Wasserstein-like distances on simplexes. The preliminary section of the paper contains all required definitions and facts from the theory of simplexes, sufficient statistics, and ergodic decompositions of measures. Describing this theory, we mostly follow [11] and [17].

In Section 3 we define the Kantorovich problem with additional restrictions. We also formulate notions of “good” (in some defined sense) linear restrictions: *weakly regular*, *ergodic decomposable*, *coherent* and *geometric* ones. These notions are important for the formulation of the main results.

In Section 4 we provide a proof of a measurable selection statement, based on the results of Rieder [22] about measurability of solutions in optimization problems.

In Section 5 we formulate and prove the main statement about existence of the decomposition of a Kantorovich problem in the case an additional restriction is ergodic decomposable, weakly regular, and coherent. In Section 6 we prove the result about decomposition of Wasserstein-like distances in the case an additional restriction is geometric (in addition to the previous assumptions).

We also describe some possible applications of the obtained results. These applications are closely related to the study of symmetric and invariant modifications of Kantorovich problem, which are studied, for example, in [4], [8], [16], [27].

2. PRELIMINARIES: SIMPLEXES AND DISINTEGRATION OF MEASURES

In this section we briefly discuss definitions and results from the theory of simplexes, ergodic decompositions and disintegrations of measures, that will be used in the rest of the paper.

The standard way to define a notion of infinite-dimensional simplex is the approach of Choquet.

Definition 2.1. A point $x \in K$ of a convex compact set K is extreme iff for any two points $a, b \in K$, such that $x = t \cdot a + (1 - t) \cdot b$ for some $t \in (0, 1)$, it follows that $a = b$. Compact convex metrizable set K is a **Choquet simplex** iff for every element $a \in K$ of this set there is a **unique** Borel measure μ_a on K such that

- $\mu_a(\partial_e(K)) = 1$, where $\partial_e(K)$ is a subset consisted of all extreme points of K ,
- $a = \text{bar}(\mu_a) := \int_K x d\mu_a$.

Affine map $T : K_1 \rightarrow K_2$ between two convex sets is a map with the property $T(\alpha \cdot a + (1 - \alpha) \cdot b) = \alpha \cdot T(a) + (1 - \alpha) \cdot T(b)$ for all $\alpha \in [0, 1]$, $a, b \in K_1$. Two simplexes are isomorphic iff there exists an affine homeomorphism between them. There is only one n -dimensional Choquet simplex up to isomorphism for any finite n .

Bauer simplex is a Choquet simplex with closed (hence compact) Choquet boundary. There are infinitely many non-isomorphic Bauer simplexes. **Poulsen simplex** is a Choquet simplex with dense Choquet boundary. There is only one (up to isomorphism) Poulsen simplex (see [21] for details).

Example 2.2. Let (X, d) be a compact metric space, $\mathcal{P}(X)$ be the set of all probability measures equipped with the weak*-topology, induced by the embedding $\mathcal{P}(X) \subset C(X)^*$. Then

$\mathcal{P}(X)$ is a Bauer simplex, its Choquet boundary consists of all Dirac measures. Every Bauer simplex is isomorphic to $\mathcal{P}(X)$ for some compact metrizable space X (see [3]).

Example 2.3. Let (X, d) be a compact metric space, G be an amenable group, acting on X by homeomorphisms, $\mathcal{P}_G(X)$ be the set of all invariant probability measures equipped with the weak*-topology, induced by the embedding $\mathcal{P}(X) \subset C(X)^*$. Then $\mathcal{P}_G(X)$ is a Choquet simplex, its Choquet boundary consists of all ergodic measures. Every Choquet simplex is isomorphic to $\mathcal{P}_G(X)$ for some compact metrizable X and some continuous action of the group \mathbb{Z} (see [9], [10]).

Example 2.4. Let \mathcal{C} be a C^* -algebra, $S(\mathcal{C})$ be its state space, i.e. set of all positive continuous linear functionals of norm one, $S(\mathcal{C})$ be equipped with weak*-topology, induced by the embedding $S(\mathcal{C}) \subset \mathcal{C}^*$. Then $S(\mathcal{C})$ is a compact convex set. It is a Choquet simplex if and only if \mathcal{C} is a commutative algebra (see [3]).

Example 2.5. Let (X, d) be a Polish metric space, $\mathcal{P}(X)$ be the set of all probability measures equipped with the weak*-topology, induced by the embedding $\mathcal{P}(X) \subset C(X)^*$. Then $\mathcal{P}(X)$ is a Choquet (and Bauer) simplex if and only if X is compact.

It can be noted, that this approach to the notion of simplex has two substantial disadvantages:

- At first, all Choquet simplexes are assumed to be compact sets by the definition. It is convenient in many cases, but spaces of measures, being equipped with Kantorovich metric, are not compact in general.
- The second, and, arguably, more important “contra” for the Choquet theory: it does not link explicitly the representation of an element of simplex as a mixture of the extreme points with an ergodic decomposition of the measure corresponding to the element.

The approach, introduced by Dynkin in [11] and developed subsequently by several authors in [17] is free of these disadvantages. The notion of simplex provided by Dynkin is a generalization of the one introduced by Choquet. It is formulated in purely measure-theoretic terms and does not require any topological assumptions. For a special subclass of Dynkin simplexes (that are called *ergodic decomposable* in [17]) there is a result that connects representation of a measure as a mixture of extremes with its disintegration w.r.t. some appropriate σ -subalgebra. Conditional measures of this disintegration correspond to the extreme points of a simplex.

Let us denote via $B(X, \mathcal{A})$ the set of all \mathcal{A} -measurable bounded real-valued functions on X .

Definition 2.6. Let (X, \mathcal{A}) be a measurable space, $M \subset \mathcal{P}(X)$ is a subset of probability measures. The smallest σ -algebra on M , such that for any $f \in B(X, \mathcal{A})$, $\mu \in M$ the map $\mu \rightarrow \mu(f)$ is measurable, is called a **standard** σ -algebra.

Definition 2.7. Let (X, \mathcal{A}) be a measurable space, $M \subseteq \mathcal{P}(X)$ be a subset equipped with a standard σ -algebra. Let us define **barycenter** $\text{bar}: \mathcal{P}(M) \rightarrow M$ by the formula

$$\text{bar}(\tilde{\mu})(f) = \int_M \left(\int_X f d\nu \right) d\tilde{\mu}(\nu), \quad \forall f \in B(X, \mathcal{A}).$$

The **boundary** of (M, \mathcal{B}) is a set $\partial_e(M) \subset M$ of such points m , that for any measure $\mu \in \mathcal{P}(M)$ the property $\text{bar}(\mu) = m$ implies $\mu_m(\{m\}) = 1$. A measurable space (M, \mathcal{B}) is called Dynkin

simplex, iff its boundary $\partial_e(M)$ is measurable, and for any $\mu \in M$ there exists a unique $\tilde{\mu} \in \mathcal{P}(M)$ s.t. $\text{bar}(\tilde{\mu}) = \mu$ and $\tilde{\mu}(\partial_e(M)) = 1$.

Example 2.8. If $M := \mathcal{P}(X)$, then it is a Dynkin simplex with boundary $\partial_e(M)$ consisted of all Dirac measures.

Example 2.9. Let (X, \mathcal{A}) be a measurable space, \mathbb{G} be an amenable group, acting on X by measurable transformations, $\mathcal{P}_{\mathbb{G}}(X)$ is a set of all invariant measures. If $M = \mathcal{P}_{\mathbb{G}}(X)$, \mathcal{B} is a σ -algebra generated by evaluations, then (M, \mathcal{B}) is a Dynkin simplex, its Dynkin boundary consists of all ergodic measures. Since any Choquet simplex is isomorphic to $\mathcal{P}_{\mathbb{G}}(X)$ for some compact metrizable X and some continuous action of the group \mathbb{Z} (see. [9], [10]), every Choquet simplex is a Dynkin simplex, up to an isomorphism.

Let (X, \mathcal{A}) be some measurable space.

Definition 2.10. A function $Q : X \times \mathcal{A} \rightarrow [0, 1]$ is called **Markov kernel** on (X, \mathcal{A}) iff for each $x \in X$, $Q^x := Q(x, \cdot)$ is a probability measure on \mathcal{A} , and for each $A \in \mathcal{A}$, $x \rightarrow Q(x, A)$ is a \mathcal{A} -measurable function.

We call two Markov kernels Q_1, Q_2 on (X, \mathcal{A}) M -equivalent for some $M \subseteq \mathcal{P}(X)$ iff $\mu(\{x : Q_1(x, \cdot) = Q_2(x, \cdot)\}) = 1$ for all $\mu \in M$.

Definition 2.11. Let $(X, \mathcal{A}), (Y, \mathcal{B})$ be two measurable spaces. A function $Q : Y \times \mathcal{A} \rightarrow [0, 1]$ is called **Markov transition kernel** from (X, \mathcal{A}) to (Y, \mathcal{B}) iff for each $y \in Y$, $Q^y := Q(y, \cdot)$ is a probability measure on (X, \mathcal{A}) , and for each $A \in \mathcal{A}$, $y \rightarrow Q(y, A)$ is a \mathcal{B} -measurable function.

We call two Markov transition kernels Q_1, Q_2 from (X, \mathcal{A}) to (Y, \mathcal{B}) M -equivalent for some $M \subseteq \mathcal{P}(Y)$ iff $\mu(\{y : Q_1(y, \cdot) = Q_2(y, \cdot)\}) = 1$ for all $\mu \in M$. Each Markov kernel Q defines a positive operator on $B(X, \mathcal{A})$ by the formula

$$Q(f)(x) := \int f dQ(x, \cdot).$$

Definition 2.12. Let $M \subseteq \mathcal{P}(X, \mathcal{A})$, \mathcal{A}^0 be a σ -subalgebra of \mathcal{A} . A Markov kernel Q on (X, \mathcal{A}) is called a **decomposition** for a triple $(\mathcal{A}, \mathcal{A}^0, M)$ iff

$$\mathbb{E}_{\mu}(f|\mathcal{A}^0) = Q(f) \text{ a.e. x w.r.t. } \mu$$

for all $f \in B(X, \mathcal{A})$, $\mu \in M$. Here $\mathbb{E}_{\mu}(f|\mathcal{A}^0)$ is a conditional expectation of f with respect to σ -algebra \mathcal{A}^0 and measure μ .

Note, that a decomposition for a triple $(\mathcal{A}, \mathcal{A}^0, M)$ is by definition a Markov transition kernel from (X, \mathcal{A}) to (X, \mathcal{A}^0) .

Define the action of Markov kernel on $\mathcal{P}(X, \mathcal{A})$ as follows:

$$Q_{\#}(\mu)(A) := \int_X Q(x, A) d\mu \quad \forall A \in \mathcal{A}.$$

A measure $\mu \in \mathcal{P}$ is said to be invariant w.r.t. Q iff $Q_{\#}(\mu) = \mu$.

Analogously, for any Markov transition kernel there is a map from $B(X, \mathcal{A})$ to $B(Y, \mathcal{B})$:

$$Q(f)(y) := \int f(x) dQ(y, \cdot)$$

and a map from $\mathcal{P}(Y, \mathcal{B})$ to $\mathcal{P}(X, \mathcal{A})$

$$Q_{\#}(\nu)(A) := \int_Y Q(y, A) d\nu \quad \forall A \in \mathcal{A}.$$

Definition 2.13. $M \subseteq \mathcal{P}(X)$ is **separable** iff \mathcal{A} contains a countable family \mathcal{F} of subsets, such that for each pair of different measures $\mu_1, \mu_2 \in M$ there exists $A \in \mathcal{F}$ s.t. $\mu_1(A) \neq \mu_2(A)$.

Definition 2.14. A σ -algebra $\mathcal{A}^0 \subseteq \mathcal{A}$ is called **sufficient** for a separable $M \subseteq \mathcal{P}(X)$, if there is a Markov kernel Q on (X, \mathcal{A}) , such that it is a decomposition for a triple $(\mathcal{A}, \mathcal{A}^0, M)$.

Definition 2.15. A sufficient σ -algebra is called **H -sufficient** for M if it is sufficient and

$$\mu(\{x : Q^x \in M\}) = 1 \quad \forall \mu \in M.$$

We call two σ -subalgebras $\mathcal{A}^1 \subseteq \mathcal{A}$, $\mathcal{A}^2 \subseteq \mathcal{A}$ M -equivalent for some $M \subseteq \mathcal{P}(X, \mathcal{A})$ iff for any $\mu \in M$, any $A_1 \in \mathcal{A}^1$ there exists $A_2 \in \mathcal{A}^2$ s.t. $A_1 = A_2$ a.e. w.r.t. μ .

The following important theorem is the union of the statements of Theorem 3.1, Theorem 3.2, Theorem 3.3 of [11] and Lemma 3.6 of [17].

Theorem 2.16. Assume that \mathcal{A} is a countably generated σ -algebra. Then for a separable set $M \subseteq \mathcal{P}(X, \mathcal{A})$, the following properties are equivalent.

- There exists a unique, up to M -equivalence, H -sufficient σ -algebra $\mathcal{A}^0 \subseteq \mathcal{A}$ for M , which is the σ -algebra of all sets $A \in \mathcal{A}$ s.t. $\mu(A) = 0$ or $\mu(A) = 1 \quad \forall \mu \in M$.
- There exists a Markov kernel Q on (X, \mathcal{A}) with the property:

$$Q(gQ(f)) = Q(g)Q(f) \quad \forall f, g \in B(X, \mathcal{A}),$$

or, equivalently, with the property

$$Q^x(\{y \in X : Q^x = Q^y\}) = 1 \quad \forall x \in X,$$

such that M is a set of all Q -invariant elements of $\mathcal{P}(X, \mathcal{A})$.

For a Markov kernel from the proposition, the corresponding H -sufficient algebra is the algebra (up to M -equivalence) of all Q -invariant measurable sets.

Definition 2.17. A separable set $M \subseteq \mathcal{P}(X, \mathcal{A})$ of measures on a countably generated σ -algebra \mathcal{A} is called **ergodic decomposable simplex** iff any of the equivalent properties of Theorem 2.16 is satisfied.

In [11] Dynkin provide a series of examples of ergodic decomposable simplexes. In particular, the simplexes from examples 2.8, 2.9 are ergodic decomposable.

The following result is a direct consequence of Theorem 3.1 from [11] and Remark 3.8 from [17].

Theorem 2.18. Ergodically decomposable simplex $M \subseteq \mathcal{P}(X, \mathcal{A})$ is a Dynkin simplex. Moreover, there exists such a Markov kernel Q from the definition of ergodic decomposable simplex, that Dynkin boundary is defined as following:

$$\partial_e(M) = \{Q^x : x \in X\}$$

and

$$\text{bar}(\tilde{\mu}) = \mu \iff \tilde{\mu}(S) = \mu(\{x : Q^x \in S\})$$

for any measurable $S \subseteq M$.

The following statement establishes a coincidence (in the case of a Polish space X) of the Borel σ -algebra generated by the topology of weak convergence on $\mathcal{P}(X)$, and the standard σ -algebra on $\mathcal{P}(X)$.

Proposition 2.19. Let (X, \mathcal{A}) be a Polish space with Borel σ -algebra. Then the following three σ -algebras on $\mathcal{P}(X)$ coincide:

- (1) the standard σ -algebra \mathcal{E} ,
- (2) σ -algebra, generated by all sets of the form $\{\mu \in \mathcal{P}(X) : a \leq \mu(A) \leq b\}$, $A \in \mathcal{A}$, $a, b \in \mathbb{Q} \cap [0, 1]$,
- (3) Borel σ -algebra, generated by the topology of weak convergence on $\mathcal{P}(X)$.

Proof. Let us show first, that \mathcal{E} can be generated by all functionals of the form $\mu \rightarrow \mu(A)$, $A \in \mathcal{A}$. It is obvious, that if all maps $\mu \rightarrow \mu(A)$ are measurable, then the map $\mu \rightarrow \mu(s) := \int s d\mu$ is measurable for every step-function s (step function is a linear combination of a finite number of measurable indicator functions). By definition of Lebesgue integration, $\mu(f) := \int f d\mu := \sup\{\mu(s) : 0 \leq s \leq f^+\} - \sup\{\mu(s) : 0 \leq s \leq f^-\}$ (where $f = f^+ - f^-$, f^+, f^- are non-negative and measurable, s is a measurable step-function) and its measurability follows from the classic fact, that a pointwise supremum of measurable maps is measurable.

Since the family of all intervals with rational endpoints in $[0, 1]$ generates Borel σ -algebra on $[0, 1]$, the standard σ -algebra coincides with the σ -algebra generated by all sets of the form $\{\mu \in \mathcal{P}(X) : a \leq \mu(A) \leq b\}$, $A \in \mathcal{A}$, $a, b \in \mathbb{Q} \cap [0, 1]$.

The equivalence of this σ -algebra with the Borel one (w.r.t. topology of weak convergence) was proved in [14] (Theorem 2.3). \square

Let us discuss two important examples of ergodic decomposable simplexes.

Example 2.20 (Main example). Let \mathbb{G} be locally compact amenable group with a fixed continuous action on a Polish metric space (X, d) . An action of \mathbb{G} on $X \times X$ is defined in the “diagonal” way: $g(x_1, x_2) := (g(x_1), g(x_2))$, where g is an action of the element $g \in \mathbb{G}$ on X . A measure μ on (X, d) is called *invariant* w.r.t. \mathbb{G} iff $\mu \circ g^{-1} = \mu$ for every $g \in \mathbb{G}$. Denote via $\mathcal{P}_{\mathbb{G}}(X) \subseteq \mathcal{P}(X)$ the set of all Borel invariant probability measures on (X, d) . It is known that $\mathcal{P}_{\mathbb{G}}(X)$ is a closed ergodic decomposable simplex (see Sections 6 and 7 in [11] for the proof). The corresponding H -sufficient σ -algebra can be defined as an algebra \mathcal{A}^{inv} of all Borel \mathbb{G} -invariant measurable subsets. The corresponding Markov kernel Q on (X, \mathcal{A}) is defined as

$$\tilde{Q}(x, A) := \lim_{n \rightarrow \infty} \frac{1}{\mu(F_n)} \int_{F_n} g_{\#}(\delta_x(A)) d\nu(g).$$

Here (F_n) is a Følner sequence for \mathbb{G} , ν is a left Haar measure on \mathbb{G} , δ_x is the Dirac measure concentrated in $x \in X$. Recall that a Følner sequence for \mathbb{G} is a sequence of nonempty compact subsets of \mathbb{G} , such that

$$\lim_{n \rightarrow \infty} \frac{\mu(gF_n \triangle F_n)}{\mu(F_n)} = 0$$

for each $g \in \mathbb{G}$, where $gF_n := \{gf : f \in F_n\}$, and \triangle is the symmetric difference.

Example 2.21 (Discrete-time Markov process). Let (X, \mathcal{A}) be a Polish space with Borel σ -algebra, $Q : X \times \mathcal{A} \rightarrow [0, 1]$ be a Markov kernel with the property $Q(fQ(g)) = Q(f)Q(g)$.

The sequence of probability measures $\mu_n := Q^n(\mu_0)$, $n \in \mathbb{N}$, is called a discrete-time Markov process with the initial distribution μ_0 and the transition kernel Q .

In this case the set of all Q -invariant probability measures on X (stationary distributions) $\mathcal{P}_Q(X) \subseteq \mathcal{P}(X)$ is an ergodic decomposable simplex. The simplex is closed if the corresponding Markov operator $Q : \mathcal{P}(X) \rightarrow \mathcal{P}(X)$ is weak*-continuous. Its H -sufficient σ -algebra is equivalent to the algebra \mathcal{A}_Q of all Borel Q -invariant subsets (i.e. such $A \in \mathcal{A}$, that $Q(I_A) = I_A$, where I_A is an indicator function of A). See Section 9 of [11] for the generalization of this example to a continuous-time case.

3. KANTOROVICH PROBLEM WITH ADDITIONAL LINEAR RESTRICTIONS

Let (X, \mathcal{A}) , (Y, \mathcal{B}) be two Polish spaces with Borel σ -algebras. Denote via $\mathcal{P}(X)$, $\mathcal{P}(Y)$, $\mathcal{P}(X \times Y)$ the sets of all probability measures on (X, \mathcal{A}) , (Y, \mathcal{B}) and $(X \times Y, \mathcal{A} \otimes \mathcal{B})$ respectively. Note that it is possible to consider sets of measures as subsets of Banach dual spaces: $\mathcal{P}(X) \subset C_b(X)^*$, $\mathcal{P}(Y) \subset C_b(Y)^*$, $\mathcal{P}(X \times Y) \subset C_b(X \times Y)^*$. Consider weak*-topologies on dual spaces and equip the spaces of probability measures with the topologies induced by the inclusion. It is known, that the defined topological spaces $\mathcal{P}(X)$, $\mathcal{P}(Y)$, $\mathcal{P}(X \times Y)$ are Polish (see [20], Theorem 6.2 and 6.5). Moreover, their topology coincides with the topology of weak convergence of measures. Due to this fact, the notions of weak and weak*-topology on the spaces of probability measures will be used interchangeably.

Let us define projection operators $\text{Pr}_X : \mathcal{P}(X \times Y) \rightarrow \mathcal{P}(X)$, $\text{Pr}_Y : \mathcal{P}(X \times Y) \rightarrow \mathcal{P}(Y)$ as follows:

$$(3.1) \quad \text{Pr}_X(\pi)(A) = \pi(A \times Y), \quad \forall A \in \mathcal{A},$$

$$(3.2) \quad \text{Pr}_Y(\pi)(B) = \pi(X \times B), \quad \forall B \in \mathcal{B}.$$

In addition, let us define $\text{Pr} : \mathcal{P}(X \times Y) \rightarrow \mathcal{P}(X) \times \mathcal{P}(Y)$ as $\text{Pr}(\pi) = (\text{Pr}_X(\pi), \text{Pr}_Y(\pi))$. It is clear, that the defined operators are weakly continuous.

In Kantorovich theory we are interested in the following sets of measures:

$$\Pi(\mu, \nu) = \{\pi \in \mathcal{P}(X \times Y) : \text{Pr}(\pi) = (\mu, \nu)\}, \quad \mu \in \mathcal{P}(X), \quad \nu \in \mathcal{P}(Y).$$

Elements of these sets are called **transport plans**. Since the map Pr is continuous, the set $\Pi(\mu, \nu)$ is closed. It is also known to be compact (see [2], [6], [26]).

Let us define **cost functional** as a functional $C : \mathcal{P}(X \times Y) \rightarrow \mathbb{R} \cup \{+\infty\}$ that is affine:

$$\alpha C(\pi) + (1 - \alpha)C(\gamma) = C(\alpha\pi + (1 - \alpha)\gamma)$$

for all $\alpha \in [0, 1]$, $\pi, \gamma \in \mathcal{P}(X \times Y)$, $0 \cdot (+\infty) := 0$. In most cases we are interested in weakly lower semi-continuous (l.s.c.) cost functionals. Examples of such cost functionals can be provided by integration of lower semi-continuous functions bounded below. If $c : X \times Y \rightarrow \mathbb{R}$ is such a function, then $\text{Cost}(\pi) := \int c d\pi$ is a cost functional. Meanwhile, the variety of cost functionals is not reduced to the functionals of this form. For example, in [24] was introduced a more general class of *virtually-continuous* functions. Integration of bounded below function of this class also defines a weakly l.s.c. functional as well.

Let us fix some cost functional C and define two sets of optimal plans:

$$\Pi^{opt}(X \times Y) := \{\pi \in \mathcal{P}(X \times Y) : C(\pi) \leq C(\gamma), \quad \forall \gamma \in \Pi(\text{Pr}(\pi))\},$$

$$\Pi^{opt}(\mu, \nu) := \{\pi \in \Pi(\mu, \nu) \cap \Pi^{opt}(X \times Y)\}.$$

In some applications the Kantorovich problem appears in a modified form. For example, one may be interested in the search of optimal element not in the set of all transport plans, but only among invariant (in some defined sense) ones. Such problem is sometimes called *invariant* or *symmetric* Kantorovich problem. Another example is martingale Kantorovich problem, where optimal solution is seek in the set of plans-martingales. A fruitful way to formalize and generalize such modifications is the notion of the Kantorovich problem with additional linear restriction, which is the main object of interest in this section.

Many aspects of the Kantorovich problem in the symmetric context were described in [16]. See also [8] for the results about symmetric Monge-Kantorovich problem in the multi-marginal setting, [4] and [27] for duality and monotonicity results about Kantorovich problem with additional linear restriction. Let us start with a formalization of the notion of additional linear restriction.

Definition 3.1. For a given pair of measurable spaces (X, \mathcal{A}) , (Y, \mathcal{B}) let us call by **linear restriction** a triple $R = (\Omega, M^X, M^Y)$, consisted of a subset of measurable functions $\Omega \subseteq L^0(X \times Y, \mathcal{A} \otimes \mathcal{B})$ and two nonempty sets of measures $M^X \subseteq \mathcal{P}(X)$, $M^Y \subseteq \mathcal{P}(Y)$.

Let us define bounded sets of transport plans as follows:

$$(3.3) \quad \Pi_R(X \times Y) := \left\{ \pi \in \mathcal{P}(X \times Y) : \int \omega d\pi = 0 \quad \forall \omega \in \Omega, \Pr_X(\pi) \in M^X, \Pr_Y(\pi) \in M^Y \right\},$$

$$(3.4) \quad \Pi_R(\mu, \nu) := \left\{ \pi \in \Pi_R(X \times Y) : \Pr_X(\pi) = \mu, \Pr_Y(\pi) = \nu \right\}.$$

Two linear restrictions R_1, R_2 are called equivalent iff $\Pi_{R_1}(X \times Y) = \Pi_{R_2}(X \times Y)$.

Everywhere in this section we assume that both (X, \mathcal{A}) , (Y, \mathcal{B}) are *Polish* spaces with *Borel* σ -algebras.

Let us formulate a property of a linear restriction, which will be used in the next section to obtain an existence result for measurable selection.

Definition 3.2. Let us call linear restriction $R = (\Omega, M^X, M^Y)$ weakly regular iff

- (1) M^X, M^Y are closed in the topology of weak convergence,
- (2) the functional $\pi \rightarrow \int \omega d\pi$ is weakly continuous on $\Pi(M^X, M^Y)$ for every $\omega \in \Omega$,
- (3) the set $\Pi_R(\mu, \nu)$ is nonempty for any $\mu \in M^X, \nu \in M^Y$.

Proposition 3.3. Let $R = (\Omega, M^X, M^Y)$ be weakly regular linear restriction. Then $\Pi_R(X \times Y)$ is closed, and $\Pi_R(\mu, \nu)$ is compact for any pair of measures $(\mu, \nu) \in M^X \times M^Y$. Both spaces, $\Pi_R(X \times Y)$ and $\Pi_R(\mu, \nu)$, appear to be Polish.

Proof. It follows from the fact, that $\Pi(\mu, \nu)$ is compact in the topology of weak convergence (see [6]). \square

Let us provide an (obvious) condition on a linear restriction R , which is sufficient for its weak regularity.

Remark 3.4. If $\Omega \subseteq C_b(X \times Y)$, $M^X \subseteq \mathcal{P}(X)$ and $M^Y \subseteq \mathcal{P}(Y)$ are weakly closed and nonempty, and $\mu \otimes \nu \in \Pi_R(\mu, \nu)$ for any pair of measures $\mu \in M^X, \nu \in M^Y$, then $R = (\Omega, M^X, M^Y)$ is a weakly regular restriction.

Let us fix, in addition to the restriction R , some cost functional C . We define two sets of restricted optimal plans:

$$(3.5) \quad \Pi_R^{opt}(X \times Y) := \left\{ \pi \in \Pi_R(X \times Y) : C(\pi) \leq C(\gamma) \ \forall \gamma \in \Pi_R(\text{Pr}(\pi)) \right\},$$

$$(3.6) \quad \Pi_R^{opt}(\mu, \nu) := \left\{ \pi \in \Pi_R(\mu, \nu) \cap \Pi_R^{opt}(X \times Y) \right\}.$$

Proposition 3.5. If $R = (\Omega, M^X, M^Y)$ is weakly regular restriction, and C is weakly l.s.c., then $\Pi_R^{opt}(\mu, \nu)$ is a compact subspace of $\mathcal{P}(X \times Y)$ (for any $\mu \in M^X, \nu \in M^Y$).

Proof. By definition of lower semi-continuity of a functional, its lower level sets are closed. Thus the set $\{\pi \in \Pi_R(\mu, \nu) : C(\pi) \leq a\}$ is closed for any $a \in \mathbb{R}$. Let $a = \inf \{C(\pi) : \pi \in \Pi_R(\mu, \nu)\}$, then the corresponding level set is $\Pi_R^{opt}(\mu, \nu)$. Since it is a closed subset of a weakly compact Polish space $\Pi_R(\mu, \nu)$ (by Proposition 3.3), it is itself weakly compact and Polish. \square

As we see in the next sections, the assumption of weak regularity is enough to prove the existence of a measurable selection of optimal transport plans in the restricted set. But for the decomposition result, analogous to Theorem 1.1, we need to assume more.

By the end of the section we shall assume, that $R := (\Omega, M^X, M^Y)$ is a linear restriction, M^X, M^Y ergodic decomposable simplexes, $\mathcal{A}^0 \subseteq \mathcal{A}, \mathcal{B}^0 \subseteq \mathcal{B}$ are correspondent H -sufficient σ -algebras.

The following property of linear restriction is the key one for the formulation of the main decomposition result.

Definition 3.6. A linear restriction $R := (\Omega, M^X, M^Y)$ is called **ergodic decomposable** linear restriction, if there exists such an ergodic decomposable simplex $M \subseteq \mathcal{P}(X \times Y)$ and its correspondent H -sufficient σ -algebra $(\mathcal{A} \otimes \mathcal{B})^0 \subseteq \mathcal{A} \otimes \mathcal{B}$, that

- $\Pi_R(X \times Y) \subseteq M$,
- $\mathcal{A}^0 \otimes \mathcal{B}^0 \subseteq (\mathcal{A} \otimes \mathcal{B})^0$,
- $\text{Pr}_X(\gamma) \in \partial_e(M^X), \text{Pr}_Y(\gamma) \in \partial_e(M^Y), \forall \gamma \in \partial_e(M)$.

Let us assume, that $R := (\Omega, M^X, M^Y)$ is an ergodic decomposable restriction, $M \subseteq \mathcal{P}(X \times Y)$ is the correspondent ergodic decomposable simplex, Q_M and $(\mathcal{A} \otimes \mathcal{B})^0 \subseteq \mathcal{A} \otimes \mathcal{B}$ are the associated Markov kernel and H -sufficient σ -algebra.

Let us introduce a property of linear restriction, which assures coherency of simplexes of marginal measures and the additional restriction on transport plans.

Definition 3.7. A linear restriction $R := (\Omega, M^X, M^Y)$ is called coherent for M , if the inclusion $\pi \in \Pi_R(X \times Y)$ implies that $\int_S \omega d\pi = 0$ for any $\omega \in \Omega, S \in (\mathcal{A} \otimes \mathcal{B})^0$.

In practice, we shall use the following sufficient condition for coherency.

Proposition 3.8. If there exists such a family $\{F_\alpha\}$ of maps $F_\alpha : B(X \times Y, \mathcal{A} \otimes \mathcal{B}) \rightarrow B(X \times Y, \mathcal{A} \otimes \mathcal{B})$, $\alpha \in \Lambda$, that

- (1) $F_\alpha(gf) = gF_\alpha(f)$ for any $g \in B(X \times Y, (\mathcal{A} \otimes \mathcal{B})^0), f \in B(X \times Y, \mathcal{A} \otimes \mathcal{B}), \alpha \in \Lambda$,
- (2) linear restrictions $R = (M^X, M^Y, \Omega_{cont})$ and $R_{meas} = (M^X, M^Y, \Omega_{meas})$ are equivalent, where $\Omega_{cont} := \text{span}\{f - F_\alpha(f) : f \in C_b(X \times Y), \alpha \in \Lambda\}$, $\Omega_{meas} := \text{span}\{f - F_\alpha(f) : f \in B(X \times Y, \mathcal{A} \otimes \mathcal{B}), \alpha \in \Lambda\}$,

then R is a coherent linear restriction.

Proof. We wish to prove that $\int I_S(f - F_\alpha(f))d\pi = 0$ for any $f \in C_b(X \times Y)$, $S \in (\mathcal{A} \otimes \mathcal{B})^0$, $\pi \in \Pi_R(X \times Y)$, $\alpha \in \Lambda$, where I_S — is the indicator function of S . Since $\Pi_R(X \times Y) = \Pi_{R_{meas}}(X \times Y)$, $\int (f - F_\alpha(f))d\pi = 0$ for any $\pi \in \Pi_R(X \times Y)$, $f \in B(X \times Y, \mathcal{A} \otimes \mathcal{B})$, $\alpha \in \Lambda$. The statement of interest follows from the fact that $\int I_S(f - F_\alpha(f))d\pi = \int (I_S f - F_\alpha(I_S f))d\pi$ and $I_S f \in B(X \times Y, \mathcal{A} \otimes \mathcal{B})$ for every $f \in C_b(X \times Y)$, $S \in (\mathcal{A} \otimes \mathcal{B})^0$, $\alpha \in \Lambda$. \square

Remark 3.9. In the case $\Pi_R(\mu, \nu) = \Pi(\mu, \nu) \cap M$, $\forall (\mu, \nu) \in M^X \times M^Y$ for some ergodic decomposable restriction R , it is true that R is coherent for M . Indeed, by definition, M is the set of all Q_M -invariant measures from $\mathcal{P}(X \times Y)$. It is enough to consider $F_1(f) := Q_M(f)$, $\Lambda = \{1\}$, $\tilde{\Omega} := \{f - Q_M(f), \forall f \in C_b(X \times Y)\}$. It follows, that the restriction $\tilde{R} := (\tilde{\Omega}, M^X, M^Y)$ is coherent and equivalent to R .

As we see later, the assumptions of coherency, ergodic decomposability, and weak regularity together imply the decomposition result for the associated Kantorovich problem.

Let $d: X \times X \rightarrow \mathbb{R}$ be a given distance function, $\text{Dom}_Q \subseteq \mathcal{P}(X)$ be an ergodic decomposable simplex, $R = (\Omega_Q, \text{Dom}_Q, \text{Dom}_Q)$ be a linear restriction. Define for each number $p \in [1, +\infty)$ a function $W_p^R: \text{Dom}_Q \times \text{Dom}_Q \rightarrow [0, \infty]$ by the formula:

$$(3.7) \quad W_p^R(\mu, \nu) := \inf \left\{ \left(\int d^p(x, y) d\pi \right)^{\frac{1}{p}} : \pi \in \Pi_R(\mu, \nu) \right\}.$$

This function does not satisfy distance axioms in general. This motivates the following, another one, assumption about liner restriction, which assures W_p^R to be a distance function.

Definition 3.10. A linear restriction $R = (\Omega_Q, \text{Dom}_Q, \text{Dom}_Q)$ is a **geometric** linear restriction for Q iff the simplex Dom_Q is weakly closed, $\Omega_Q \subseteq C_b(X \times Y)$, and for every $\omega \in \Omega_Q$ it is true that:

- (1) $((Id, Id)_\# \mu)(\omega) = 0 \quad \forall \mu \in \text{Dom}_Q$,
- (2) $(\mu \otimes \nu)(\omega) = 0 \quad \forall \mu, \nu \in \text{Dom}_Q$,
- (3) $\pi(\omega) = 0 \implies \pi^T(\omega) = 0 \quad \forall \pi \in \Pi(\mu, \nu), \forall \mu, \nu \in \text{Dom}_Q$, where π^T is defined as follows:

$$\int f(x, y) d\pi^T = \int f(y, x) d\pi \quad \forall f \in C_b(X \times Y).$$

Proposition 3.11. If R is a geometric linear restriction, then W_p^R is an extended distance function.

Proof. Since $\pi = (Id, Id)_\# \mu \in \Pi_R(\mu, \mu)$, $\int d^p d\pi = 0$ for this π , and $W_p^R(\mu, \mu) = 0$. By the same reason ($d^p = 0$ only on the diagonal $\{(x, x)\}$, and only plans of the form $(Id, Id)_\# \mu$ are concentrated on it) $W_p^R(\mu, \nu) \neq 0$ if $\mu \neq \nu$. Function W_p^R is symmetric, because the inclusion $\pi \in \Pi_R(\mu, \nu)$ implies $\pi^T \in \Pi_R(\nu, \mu)$ and $\int d^p d\pi = \int d^p d\pi^T$. Triangle inequality can be proved using the standard technique (see, for example, [2], Theorem 2.2) with the use of the special version of the “gluing” lemma, that is formulated below. \square

Lemma 3.12 (version of gluing). Let $R = (\Omega_Q, \text{Dom}_Q, \text{Dom}_Q)$ be a geometric linear restriction. Then for every measure $\mu_1, \mu_2, \mu_3 \in \text{Dom}_Q$, $\pi_{12} \in \Pi_R(\mu_1, \mu_2)$, $\pi_{23} \in \Pi_R(\mu_2, \mu_3)$ there is a measure $\gamma \in \mathcal{P}(X \times X \times X)$, such that $(\text{Pr}_{12})_\#(\gamma) = \pi_{12}$, $(\text{Pr}_{23})_\#(\gamma) = \pi_{23}$, $(\text{Pr}_{13})_\#(\gamma) \in \Pi_R(\mu_1, \mu_3)$.

Proof. Let us modify a proof of the standard gluing lemma (see Lemma 2.2. of [7]). Define subspace $V \subset C_b(X \times X \times X)$ as follows:

$$V := \left\{ f_{12}(x_1, x_2) + f_{23}(x_2, x_3) + \omega_{13}(x_1, x_3) : f_{12}, f_{23} \in C_b(X \times X), \omega_{13} \in \Omega_Q \right\}.$$

Let $F(f_{12} + f_{23} + \omega_{13}) := \pi_{12}(f_{12}) + \pi_{23}(f_{23})$. Check that F is well-defined on V . Consider two representations of some element of V : $f_{12} + f_{23} + \omega_{13} = \tilde{f}_{12} + \tilde{f}_{23} + \tilde{\omega}_{13}$. Note, that $\omega_{13}(x_1, x_3) - \tilde{\omega}_{13}(x_1, x_3) = \omega_1(x_1) + \omega_3(x_3)$ for some $\omega_1, \omega_3 \in \Omega_Q$. Then $f_{12} - \tilde{f}_{12} + \omega_1 = \tilde{f}_{23} - f_{23} - \omega_3$, and therefore both parts of the equality depends only on x_2 . Since

$$\pi_{12}(f_{12} - \tilde{f}_{12} + \omega_1) = \mu_2(f_{12} - \tilde{f}_{12}) = \mu_2(\tilde{f}_{23} - f_{23}) = \pi_{23}(\tilde{f}_{23} - f_{23} - \omega_3),$$

the map F is well-defined on V . It is easy to check, that F is a bounded positive linear functional. The correspondent version of Hahn-Banach theorem (Theorem 1.25 of [1]) states, that such a functional can be extended to a positive bounded functional on the entire $C_b(X \times X \times X)$. Since on the subspaces $C_b(X_k)$ value of F coincides with the integration w.r.t. measures μ_k , one can apply Rietz theorem (Theorem 7.10.6 in [5]) to the extension of F . The resulting measure will satisfy all the required properties from the statement of the lemma we are proving. \square

Proposition 3.13. Geometric linear restriction is weakly regular.

Proof. Since $(\mu \otimes \nu)(\omega) = 0$, $\mu \otimes \nu \in \Pi_R(\mu, \nu)$ for any $\mu, \nu \in \text{Dom}_Q$, Dom_Q is weakly closed and $\Omega \subseteq C_b(X \times Y)$, it follows that $R = (\Omega, \text{Dom}_Q, \text{Dom}_Q)$ is weakly regular (see Remark 3.4). \square

Let us check that for our main examples (2.20 and 3.14 of invariant and stationary measures respectively) it is possible to define additional linear restrictions, such that all (or most of) properties introduced above are satisfied.

Example 3.14 (Main example). Let us consider the linear restriction: $R = (\Omega, \text{Dom}, \text{Dom})$, where Dom is a simplex of all invariant measures as in Example 2.20,

$$\Omega := \text{span}(\{f - f \circ g : \forall f \in C_b(X \times X), \forall g \in \mathbb{G}\}).$$

It corresponds to a restriction on transport plans to be invariant w.r.t. a diagonal action of \mathbb{G} on $X \times X$ (see [27], Prop. 5.1 for the proof): $g(x, y) = (g(x), g(y))$.

Let $M = \mathcal{P}_{\mathbb{G}}(X \times Y)$ be a simplex of all invariant measures on $X \times Y$ w.r.t. this action, $(\mathcal{A} \otimes \mathcal{B})^{\text{inv}}$ be a σ -algebra of all invariant Borel subsets on $X \times Y$. It is clear, that $\Pi_R(X \times Y) \subseteq M$, $\mathcal{A}^{\text{inv}} \otimes \mathcal{B}^{\text{inv}} \subseteq (\mathcal{A} \otimes \mathcal{B})^{\text{inv}}$ and $(\mathcal{A} \otimes \mathcal{B})^{\text{inv}}$ is the H -sufficient subalgebra associated to M . To show *ergodic decomposability* of R (Definition 3.6), we need to check, that every ergodic measure from $P_{\mathbb{G}}(X \times Y)$ has ergodic marginals, which belong to $\mathcal{P}_{\mathbb{G}}(X)$. It can be easily shown by contradiction.

Let us prove that the restriction is *geometric* (Definition 3.10):

- (1) the set $\mathcal{P}_{\mathbb{G}}(X)$ is closed and $\Omega \subset C_b(X \times X)$,
- (2) $(Id, Id)_{\#} \mu(f(x, y) - f(g(x), g(y))) = \mu(f(x, x) - f(g(x), g(x))) = \mu(f(x, x)) - g_{\#} \mu(f(x, x)) = 0$ if $\mu \in \text{Dom} = \mathcal{P}_{\mathbb{G}}(X)$, $f \in C_b(X \times X)$, $g \in \mathbb{G}$, $\mu(f) := \int f d\mu$,

(3) note that

$$\begin{aligned} (\mu \otimes \nu)(f(x, y) - f(g(x), g(y))) &= \int \left(\int f(x, y) - f(g(x), y) d\mu(x) \right) d\nu(y) + \\ &+ \int \left(\int f(g(x), y) - f(g(x), g(y)) d\nu(y) \right) d\mu(x) = 0, \quad \forall \mu, \nu \in \text{Dom} = \mathcal{P}_{\mathbb{G}}(X), \quad f \in C_b(X \times X), \quad g \in \mathbb{G}, \end{aligned}$$

(4) since the inclusion $f(x, y) - f(g(x), g(y)) \in \Omega$ implies $f(y, x) - f(g(y), g(x)) \in \Omega$ for any $f \in C_b(X \times X)$, the last requirement from the definition of geometricity is satisfied.

We also need to check, that the restriction is *coherent* w.r.t. M (Definition 3.7). Note, that $\Pi_R(\mu, \nu) = \Pi(\mu, \nu) \cap \mathcal{P}(X \times Y)$, $\forall (\mu, \nu) \in M^X \times M^Y$, and use Remark 3.9.

Remark 3.15. For the example with a simplex of invariant measures Dom (Example 2.20) it is also possible to consider other meaningful and well-behaved linear restrictions. Assume that there is a given action of group \mathbb{G} on X . One can consider the action of direct product of groups $\mathbb{G} \oplus \mathbb{G}$ of $X \times X$ defined in the natural way: $(g_1, g_2)(x_1, x_2) := (g_1(x_1), g_2(x_2))$.

Let us fix a subgroup $\mathbb{H} \subseteq \mathbb{G} \oplus \mathbb{G}$ with the induced action on $X \times X$. If the projections of the subgroup on the first and second components coincide with \mathbb{G} , the associated restriction $R = (\Omega, \text{Dom}, \text{Dom})$ with

$$\Omega := \text{span}(\{f - f \circ h : \forall f \in C_b(X \times X), \forall h \in \mathbb{H}\}).$$

has the properties of ergodic decomposability and coherency, defined in this section. It can be checked in the same way we do for the diagonal action of a group.

Example 3.16 (Discrete-time Markov process). Let simplex $\text{Dom} = \mathcal{P}_Q(X)$ be defines as in Example 2.21, and let \mathcal{A}^0 be its associated σ -subalgebra. Consider Markov transition kernel Q_M from $\mathcal{A} \otimes \mathcal{A}$ to $\mathcal{A}^0 \otimes \mathcal{A}^0$ that is defined by the formula

$$Q_M(f)(x, y) := \int f(\tilde{x}, \tilde{y}) dQ^x(\tilde{x}) \otimes Q^y(\tilde{y}).$$

It can be checked, that $Q_M(fQ_M(g)) = Q_M(f)Q_M(g)$, and, therefore, by Theorem 2.16 there is an associated ergodic decomposable simplex M . Let us consider the following linear restriction $R = (\Omega, \text{Dom}, \text{Dom})$,

$$\Omega := \text{span}(\{f - Q_M(f) : \forall f \in C_b(X \times X)\}).$$

It is ergodic decomposable and coherent w.r.t. M (the arguments are analogous to the ones from the previous example).

4. MEASURABLE SELECTION OF OPTIMAL TRANSPORT PLANS

The goal of this section is to prove the existence of a measurable map $f : M^X \times M^Y \rightarrow \Pi_R^{opt}(X \times Y)$, such that $\text{Pr}(f(\mu, \nu)) = (\mu, \nu)$ (recall, that $\text{Pr}(\pi) := (\text{Pr}_X(\pi), \text{Pr}_Y(\pi))$), under the assumptions of weak regularity of R and lower semi-continuity of cost functional. Existence of such a map is required in the proof of the decomposition result, which we shall formulate in the next section, and it seems to be of interest itself.

In the Kantorovich problem without additional linear restriction this fact is well-known (see Corollary 5.22 of [26]). Its proof relies on the sufficiency result: c -cyclical monotonicity of a support of a transport plan implies its optimality. There is no known analogue of this result

in the case of the problem with additional restriction. Thus we need to invent a more direct way to prove measurability.

Our main tool is the following theorem of Rieder:

Theorem 4.1 (Rieder [22], Th. 4.1, Cor. 4.3). Let (Θ, \mathcal{A}) , (S, \mathcal{B}) be Polish spaces with Borel σ -algebras, $D \subseteq \Theta \times S$, $u : D \rightarrow \mathbb{R} \cup \{\pm\infty\}$,

$$L_c := \{(\theta, s) \in D : u(\theta, s) \leq c\}, \quad c \in \mathbb{R},$$

$L_{+\infty} := D$. If $\forall c \in (-\infty, +\infty]$, $L_c \in \mathcal{A} \otimes \mathcal{B}$, and $\forall c \in \mathbb{R}$

$$L_c(\theta) := L_c \cap \{(\theta, s) : s \in S\}$$

is compact for any $\theta \in \text{Pr}_\Theta(D)$, then there exists a measurable function $f : \text{Pr}_\Theta(D) \rightarrow S$, such that

$$u(\theta, f(\theta)) = \inf_{s \in D(\theta)} u(\theta, s), \quad \forall \theta \in \text{Pr}_\Theta(D)$$

where $D(\theta) := \text{Pr}_\Theta^{-1}(\theta) \cap D$.

Let $\Theta := M^X \times M^Y$, $S := \Pi_R(X \times Y)$, be equipped with the topology of weak convergence and corresponding Borel σ -algebras. Both Θ and S are Polish spaces, and due to weak regularity of R , S is closed. Let $D := \{(\mu, \nu, \pi) : \pi \in \Pi_R(\mu, \nu), (\mu, \nu) \in M^X \times M^Y\}$, $u((\mu, \nu), \pi) := \text{Cost}(\pi)$. Then

$$L_c = \{(\mu, \nu, \pi) : \text{Cost}(\pi) \leq c, \pi \in \Pi_R(\mu, \nu), (\mu, \nu) \in M^X \times M^Y\}, \quad c \in (-\infty, +\infty],$$

$$L_c(\mu, \nu) = \{\pi : \text{Cost}(\pi) \leq c, \pi \in \Pi_R(\mu, \nu)\}, \quad c \in \mathbb{R}$$

To apply the theorem above, we have to prove that L_c is Borel, and that $L_c(\mu, \nu)$ is compact.

We are going to use the following proposition (Himmelberg [15], Theorem 3.5).

Proposition 4.2. Let (Θ, \mathcal{A}) , (Y, \mathcal{B}) be separable metrizable spaces with Borel σ -algebras. Let $T : \Theta \rightarrow 2^S$ be a map with values in closed subsets of S , such that for every closed subset $V \subseteq Y$

$$\{\theta : T(\theta) \cap V \neq \emptyset\} \in \mathcal{A}$$

Then the graph of T :

$$\{(\theta, s) : \theta \in \Theta, s \in T(\theta)\}$$

is in $\mathcal{A} \otimes \mathcal{B}$.

Lemma 4.3. If R is weakly regular and $\text{Cost} : \Pi_R(X \times Y) \rightarrow \mathbb{R}$ is l.s.c., then the set

$$L_c := \{(\mu, \nu, \pi) : \text{Cost}(\pi) \leq c, \pi \in \Pi_R(\mu, \nu), (\mu, \nu) \in M^X \times M^Y\}, \quad c \in (-\infty, +\infty],$$

is an element of σ -algebra $\text{Bor}(M^X \times M^Y) \otimes \text{Bor}(\Pi_R(X \times Y))$.

Proof. By definition of weak regularity, $\Pi_R(X \times Y)$ is a Polish space, $\Pi_R(\mu, \nu)$ is Polish and compact. Since Cost is an l.s.c. functional, its lower level sets $\Pi_c := \{\pi \in \mathcal{P}(X \times Y) : \text{Cost}(\pi) \leq c\}$ are closed. Since $L_c(\mu, \nu) = \Pi_R(\mu, \nu) \cap \Pi_c$, $L_c(\mu, \nu)$ is compact. It should be noted, that $L_c(\mu, \nu)$ can be empty, but empty set is compact, hence there is no contradiction here.

Let $(\Omega, \mathcal{A}) := (M^X \times M^Y, \text{Bor}(M^X \times M^Y))$, $(S, \mathcal{B}) := (\Pi_R(X \times Y), \text{Bor}(\Pi_R(X \times Y)))$, $T : M^X \times M^Y \rightarrow 2^{\Pi_R(X \times Y)}$, $T(\mu, \nu) = L_c(\mu, \nu)$. Then $L_c = \{(\mu, \nu, \pi) : \pi \in L_c(\mu, \nu), (\mu, \nu) \in M^X \times M^Y\}$ is the graph of T . Note, that T has compact (hence closed) values.

Fix an arbitrary closed set $V \subseteq \Pi_R(X \times Y)$. Let us show that the set

$$T^{-1}(V) := \{(\mu, \nu) : L_c(\mu, \nu) \cap V \neq \emptyset\}$$

is closed, hence Borel. If the set is empty, it is trivially true. Assume it is not. Let $(\mu_n, \nu_n) \in T^{-1}(V)$, $n \in \mathbb{N}$, be a sequence converging to (μ, ν) in $M^X \times M^Y$. Then $\mu_n \rightarrow \mu$, $\nu_n \rightarrow \nu$ weakly, and the sequences (μ_n) , (ν_n) are tight. But for each (μ_n, ν_n) there exists a $\pi_n \in L_c(\mu_n, \nu_n) \cap V$. It follows, that $\text{Pr}(\pi_n) = (\mu_n, \nu_n)$, and the sequence (π_n) is tight (by considering products of compact sets). Let (π_{n_k}) be a weakly convergent subsequence with limit π . Since V is closed, it is sequentially closed, and $\pi \in V$. Due to weak continuity of Pr and weak l.s.c. of Cost , it is clear, that $\text{Pr}(\pi) = (\mu, \nu)$, $\text{Cost}(\pi) \leq c$. It follows by weak regularity of R , that the functional $\pi \rightarrow \int \omega d\pi$ is weakly continuous for any $\omega \in \Omega$. Hence $\int \omega d\pi = 0$, and $\pi \in L_c(\mu, \nu)$.

We obtain, that for (μ, ν) there is a $\pi \in L_c(\mu, \nu) \cap V$, hence $(\mu, \nu) \in T^{-1}(V)$, and $T^{-1}(V)$ sequentially closed. Since $M^X \times M^Y$ is metrizable, sequential closeness implies closeness, and $T^{-1}(V)$ is closed, hence Borel. By Proposition 4.2, the graph of T , which coincides with L_c , is a Borel set. It concludes the proof. \square

We are ready to formulate and prove the main result of this section.

Theorem 4.4. Let X, Y be Polish spaces with Borel σ -algebras. If $R = (\Omega, M^X, M^Y)$ is a weakly regular linear restriction, Cost is a weakly l.s.c. cost functional, then there exists a measurable map $f : M^X \times M^Y \rightarrow \Pi_R^{\text{opt}}(X \times Y)$, such that $\text{Pr}(f(\mu, \nu)) = (\mu, \nu)$.

Proof. Let us apply Theorem 4.1. Recall, that $\Theta := M^X \times M^Y$, $S := \Pi_R(X \times Y)$ are Polish spaces with Borel σ -algebras, $D := \{(\mu, \nu, \pi) : \pi \in \Pi_R(\mu, \nu), (\mu, \nu) \in M^X \times M^Y\}$, $u((\mu, \nu), \pi) := \text{Cost}(\pi)$, $L_c = \{(\mu, \nu, \pi) : \text{Cost}(\pi) \leq c, \pi \in \Pi_R(\mu, \nu), (\mu, \nu) \in M^X \times M^Y\}$ is Borel for every $c \in (-\infty, +\infty]$ by Lemma 4.3, $L_c(\mu, \nu) = \{\pi : \text{Cost}(\pi) \leq c, \pi \in \Pi_R(\mu, \nu)\}$ is compact for every $c \in \mathbb{R}$ by weak regularity of R and lower semi-continuity of Cost . It follows directly from the application of Theorem 4.1, that there exists a Borel function $f : M^X \times M^Y \rightarrow \Pi_R(X \times Y)$, such that

$$\text{Cost}(f(\mu, \nu)) = \inf_{\pi \in \Pi_R(\mu, \nu)} \text{Cost}(\pi),$$

which implies, that $f(\mu, \nu) \in \Pi_R^{\text{opt}}(\mu, \nu)$ for any $(\mu, \nu) \in M^X \times M^Y$. \square

5. DECOMPOSITION OF KANTOROVICH PROBLEM

In this section we formulate and prove an analogue of Theorem 1.1 for the restricted Kantorovich problem under the assumptions of weak regularity, ergodic decomposability, and coherency of linear restrictions.

Let (X, \mathcal{A}) , (Y, \mathcal{B}) be two Polish spaces with Borel σ -algebras, $M^X \subseteq \mathcal{P}(X)$ and $M^Y \subseteq \mathcal{P}(Y)$ be two closed ergodic decomposable simplexes. Denote via Q_X and Q_Y the Markov kernels on (X, \mathcal{A}) , (Y, \mathcal{B}) , associated with M^X and M^Y , such that $\partial_e(M^X) = \{Q_X^x\}$, $\partial_e(M^Y) = \{Q_Y^y\}$. If $\mathcal{A}^0 \subseteq \mathcal{A}$ and $\mathcal{B}^0 \subseteq \mathcal{B}$ be their corresponding H -sufficient subalgebras, then, by definition, Q_X and Q_Y are decompositions for triples $(\mathcal{A}, \mathcal{A}^0, M^X)$ and $(\mathcal{B}, \mathcal{B}^0, M^Y)$ respectively. We shall use the notation: $\xi(x) := Q_X^x$, $\eta(y) := Q_Y^y$. Note, that the maps $\xi : X \rightarrow \partial_e(M^X)$, $\eta : Y \rightarrow \partial_e(M^Y)$ are \mathcal{A}^0 - and \mathcal{B}^0 -measurable respectively.

Let $R = (\Omega, M^X, M^Y)$ be *weakly regular, ergodic decomposable* and *coherent* w.r.t. M linear restriction, where $M \subseteq \mathcal{P}(X \times Y)$ is the corresponding ergodic decomposable simplex, Q_M and $(\mathcal{A} \otimes \mathcal{B})^0 \subseteq \mathcal{A} \otimes \mathcal{B}$ are the associated Markov kernel and H -sufficient σ -algebra.

Definition 5.1. Denote by $\tilde{\Pi}(\tilde{\mu}, \tilde{\nu})$ the set of all probability measures $\tilde{\pi} \in \mathcal{P}(X \times Y, (\mathcal{A} \otimes \mathcal{B})^0)$, such that $\text{Pr}_X(\tilde{\pi}) = \tilde{\mu}$, $\text{Pr}_Y(\tilde{\pi}) = \tilde{\nu}$.

Definition 5.2. Let $\mu \in M^X$, $\nu \in M^Y$. Define the set $\Theta(R, \mu, \nu)$ as the set of all pairs (π, Q_π) , where $\pi \in \Pi_R(\mu, \nu)$, and Q_π is such a Markov transition kernel from $(X \times Y, \mathcal{A} \otimes \mathcal{B})$ to $(X \times Y, (\mathcal{A} \otimes \mathcal{B})^0)$, that $Q_\pi(\tilde{\pi}) = \pi$ and $Q_\pi^{(x,y)} \in \Pi_R(Q_X^x, Q_Y^y)$ for $\tilde{\pi}$ -a.e. $(x, y) \in X \times Y$. Here $\tilde{\pi}$ is the restriction of measure π from $\mathcal{A} \otimes \mathcal{B}$ to $(\mathcal{A} \otimes \mathcal{B})^0$.

In the following Lemma we use properties of ergodic decomposability and coherency of the linear restriction.

Lemma 5.3. Under the assumptions made above, for any $\pi \in \Pi_R(\mu, \nu)$ there exists a Markov transition kernel Q_π , such that $(\pi, Q_\pi) \in \Theta(R, \mu, \nu)$.

Proof. For each measure $\pi \in \Pi(\mu, \nu)$ define measure $\tilde{\pi}$ as the restriction of π on $(\mathcal{A} \otimes \mathcal{B})^0$. Its marginals, $\text{Pr}_X(\tilde{\pi})$ and $\text{Pr}_Y(\tilde{\pi})$, are the restrictions of μ and ν to subalgebras \mathcal{A}^0 and \mathcal{B}^0 respectively. Hence, the inclusion $\pi \in \Pi(\mu, \nu)$ implies that $\tilde{\pi} \in \tilde{\Pi}(\tilde{\mu}, \tilde{\nu})$ (see Definition 5.1).

Let $Q_\pi := Q_M$, where Q_M is the Markov kernel from the definition of ergodic decomposability of restriction R . This property implies $Q_\pi(\tilde{\pi}) = \pi$ and $Q_\pi^{(x,y)} \in \mathcal{P}(X \times Y)$. Moreover, $\text{Pr}(Q_\pi^{(x,y)}) = (\xi(x), \eta(y))$ for $\tilde{\pi}$ -a.e. (x, y) .

Let us check that $\tilde{\pi}$ -a.e. $Q_\pi^{(x,y)}(\omega) = 0$ for each function $\omega \in \Omega$. Let us use the following equality:

$$(5.1) \quad \int h dQ_\pi^{(x,y)} = \mathbb{E}_\pi(h | (\mathcal{A} \otimes \mathcal{B})^0) \quad \pi\text{-a.e. } \forall h \in B(X \times Y, \mathcal{A} \otimes \mathcal{B}).$$

By coherency of R and the definition of conditional expectation, we obtain:

$$\int_S \left(\int \omega dQ_\pi^{(x,y)} \right) d\tilde{\pi} = \int_S \omega d\pi = 0, \quad \forall S \in (\mathcal{A} \otimes \mathcal{B})^0.$$

Hence, $\int \omega dQ_\pi^{(x,y)} = 0$ for $\tilde{\pi}$ -a.e. (x, y) .

As we just proved, $Q_\pi^{(x,y)} \in \Pi_R(\xi(x), \eta(y))$ for $\tilde{\pi}$ -a.e. (x, y) . Since the map $(x, y) \rightarrow Q_\pi^{(x,y)}$ is $(\mathcal{A} \otimes \mathcal{B})^0$ -measurable, Q_π is a Markov transition kernel from $(X \times Y, \mathcal{A} \otimes \mathcal{B})$ to $(X \times Y, (\mathcal{A} \otimes \mathcal{B})^0)$.

It follows from (5.1), that

$$\int Q_\pi(f) d\tilde{\pi} = \int f d\pi, \quad \forall f \in B(X \times Y, \mathcal{A} \otimes \mathcal{B}),$$

which implies $(\pi, Q_\pi) \in \Theta(R, \mu, \nu)$ and concludes the proof. \square

Let $c: X \times Y \rightarrow \mathbb{R}$ be such a measurable function, that the functional $Cost: \mathcal{P}(X \times Y) \rightarrow \mathbb{R} \cup \{+\infty\}$, being defined as $Cost(\pi) = \int c d\pi$, appears to be lower semi-continuous w.r.t. topology of weak convergence.

Lemma 5.4. Under the assumptions made above, there exists a Markov transition kernel Q_{opt} from $(X \times Y, \mathcal{A} \otimes \mathcal{B})$ to $(X \times Y, (\mathcal{A} \otimes \mathcal{B})^0)$, such that

- (1) $\int cdQ_{opt}^{(x,y)} = \inf \left\{ \int cd\pi : \pi \in \Pi_R(\xi(x), \eta(y)) \right\} \quad \forall (x, y) \in X \times Y,$
- (2) for any pair $(\pi, Q_\pi) \in \Theta(R, \mu, \nu)$, $\int cdQ_{opt}(\tilde{\pi}) \leq \int cdQ_\pi(\tilde{\pi})$, where $\tilde{\pi}$ is the restriction of π from $\mathcal{A} \otimes \mathcal{B}$ to $(\mathcal{A} \otimes \mathcal{B})^0$,
- (3) for each measure $\tilde{\pi} \in \tilde{\Pi}(\tilde{\mu}, \tilde{\nu})$, $(Q_{opt}(\tilde{\pi}), Q_{opt}) \in \Theta(R, \mu, \nu)$, where $\tilde{\mu}$ is the restriction of μ from \mathcal{A} to \mathcal{A}^0 , $\tilde{\nu}$ is the restriction of ν from \mathcal{B} to \mathcal{B}^0 .

Proof. By Theorem 4.4, there exists a measurable map $f: M^X \times M^Y \rightarrow \Pi_R^{opt}(X \times Y)$, such that $\Pr(f(\mu, \nu)) = (\mu, \nu)$. Let $Q_{opt}^{(x,y)} := f(\xi(x), \eta(y))$. Since $x \rightarrow \xi(x)$ and $y \rightarrow \eta(y)$ are measurable maps w.r.t. algebras \mathcal{A}^0 and \mathcal{B}^0 respectively, Q_{opt} is in fact a Markov transition kernel from $(X \times Y, \mathcal{A} \otimes \mathcal{B})$ to $(X \times Y, \mathcal{A}^0 \otimes \mathcal{B}^0)$ and, consequently, a Markov transition kernel to $(\mathcal{A} \otimes \mathcal{B})^0$. It follows by definition, that

$$\int cdQ_{opt}^{(x,y)} = \inf \left\{ \int cd\gamma : \gamma \in \Pi_R(\xi(x), \eta(y)) \right\}.$$

Let $(\pi, Q_\pi) \in \Theta(R, \mu, \nu)$. Then $\int cdQ_\pi(\tilde{\pi}) := \int \left(\int cdQ_\pi^{(x,y)} \right) d\tilde{\pi}$. Since $Q_\pi^{(x,y)} \in \Pi_R(\xi(x), \eta(y))$ $\tilde{\pi}$ -a.e., $\int cdQ_\pi^{(x,y)} \geq \int cdQ_{opt}^{(x,y)} \tilde{\pi}$ -a.e., which implies $\int cdQ_\pi(\tilde{\pi}) \geq \int cdQ_{opt}(\tilde{\pi})$.

Let $\tilde{\pi} \in \tilde{\Pi}(\tilde{\mu}, \tilde{\nu})$, where $\tilde{\mu}, \tilde{\nu}$ are the restrictions of measures $\mu \in M^X, \nu \in M^Y$ to $\mathcal{A}^0, \mathcal{B}^0$ respectively. By definition of Markov transition kernel, $Q_{opt}(\tilde{\pi}) \in \mathcal{P}(X \times Y)$. Let us check that $Q_{opt}(\tilde{\pi}) \in \Pi_R(\mu, \nu)$. Since $Q_{opt}^{(x,y)} \in \Pi_R(\xi(x), \eta(y))$ for every pair $(x, y) \in X \times Y$, it can be shown, that $\Pr(Q_{opt}(\tilde{\pi})) = (\mu, \nu)$. Let us provide an argument for the first marginal:

$$\begin{aligned} (5.2) \quad \int \int f(\tilde{x}) dQ_{opt}^{(x,y)}(\tilde{x}, \tilde{y}) d\tilde{\pi}(x, y) &= \int \int f(\tilde{x}) dQ_X^x(\tilde{x}) d\tilde{\pi}(x, y) = \\ &= \int \int f(\tilde{x}) dQ_X^x(\tilde{x}) d\tilde{\mu}(x) = \int f(x) d\mu, \quad \forall f \in B(X, \mathcal{A}), \end{aligned}$$

where the last equality follows from the fact, that Q_X is a decomposition for $(\mathcal{A}, \mathcal{A}^0, M^X)$. Analogously, it can be checked for the second marginal. It is clear, that if $\int \omega dQ_{opt}^{(x,y)} = 0$ for any (x, y) , it implies the equality: $\int \omega dQ_{opt}^{(x,y)} d\tilde{\pi} = 0$. Thus $Q_{opt}(\tilde{\pi}) \in \Pi_R(\mu, \nu)$, and, hence, there exists a measure $\pi \in \Pi_R(\mu, \nu)$, such that $\pi = Q_{opt}(\tilde{\pi}) \in \Pi_R(\mu, \nu)$. \square

Let us formulate and prove the main result.

Theorem 5.5 (Main theorem). Let $(X, \mathcal{A}), (Y, \mathcal{B})$ be two Polish spaces with Borel σ -algebras; $c: X \times Y \rightarrow \mathbb{R}$ be such a measurable function, that functional $Cost: \mathcal{P}(X \times Y) \rightarrow \mathbb{R} \cup \{+\infty\}$, defined as $Cost(\pi) = \int cd\pi$, appears to be a lower semi-continuous functional w.r.t. topology of weak convergence; M^X, M^Y be two ergodic decomposable simplexes, $\mathcal{A}^0 \subseteq \mathcal{A}, \mathcal{B}^0 \subseteq \mathcal{B}$ be corresponding σ -subalgebras; $R = (\Omega, M^X, M^Y)$ be a *weakly regular, ergodic decomposable, and coherent* w.r.t. M linear restriction, where $M \subseteq \mathcal{P}(X \times Y)$ is the associated ergodic decomposable simplex, Q_M and $(\mathcal{A} \otimes \mathcal{B})^0 \subseteq \mathcal{A} \otimes \mathcal{B}$ are the associated Markov kernel and

H -sufficient σ -algebra. Then

$$(5.3) \quad \inf \left\{ \int cd\pi : \pi \in \Pi_R(\mu, \nu) \right\} = \inf \left\{ \int Q_{opt}(c)d\tilde{\pi} : \tilde{\pi} \in \tilde{\Pi}(\tilde{\mu}, \tilde{\nu}) \right\} = \\ = \inf \left\{ \int \inf \left\{ \int cd\pi : \pi \in \Pi_R(\xi(x), \eta(y)) \right\} d\tilde{\pi} : \tilde{\pi} \in \tilde{\Pi}(\tilde{\mu}, \tilde{\nu}) \right\}.$$

Proof. Let

$$\Theta_{opt}(R, \mu, \nu) := \{(Q_{opt}(\tilde{\pi}), Q_{opt}) : \tilde{\pi} \in \tilde{\Pi}(\tilde{\mu}, \tilde{\nu})\},$$

where Q_{opt} is as in Lemma 5.4. By this Lemma, $\Theta_{opt}(R, \mu, \nu) \subseteq \Theta(R, \mu, \nu)$, thus,

$$\inf \left\{ \int Q_{\pi}(c)d\tilde{\pi} : (\pi, Q_{\pi}) \in \Theta(R, \mu, \nu) \right\} \leq \inf \left\{ \int Q_{opt}(c)d\tilde{\pi} : \tilde{\pi} \in \tilde{\Pi}(\tilde{\mu}, \tilde{\nu}) \right\}.$$

As follows from Lemma 5.3,

$$\inf \left\{ \int cd\pi : \pi \in \Pi_R(\mu, \nu) \right\} = \inf \left\{ \int Q_{\pi}(c)d\tilde{\pi} : (\pi, Q_{\pi}) \in \Theta(R, \mu, \nu) \right\}.$$

By Lemma 5.4, $\int cdQ_{opt}(\tilde{\pi}) \leq \int cdQ_{\pi}(\tilde{\pi})$ for all $(\pi, Q_{\pi}) \in \Theta(R, \mu, \nu)$. Hence,

$$\inf \left\{ \int Q_{\pi}(c)d\tilde{\pi} : (\pi, Q_{\pi}) \in \Theta(R, \mu, \nu) \right\} \geq \inf \left\{ \int Q_{opt}(c)d\tilde{\pi} : \tilde{\pi} \in \tilde{\Pi}(\tilde{\mu}, \tilde{\nu}) \right\}.$$

Thus, we conclude

$$\inf \left\{ \int cd\pi : \pi \in \Pi_R(\mu, \nu) \right\} = \inf \left\{ \int Q_{opt}(c)d\tilde{\pi} : \tilde{\pi} \in \tilde{\Pi}(\tilde{\mu}, \tilde{\nu}) \right\}.$$

The equality

$$\inf \left\{ \int cd\pi : \pi \in \Pi_R(\mu, \nu) \right\} = \inf \left\{ \int \inf \left\{ \int cd\pi : \pi \in \Pi_R(\xi(x), \eta(y)) \right\} d\tilde{\pi} : \tilde{\pi} \in \tilde{\Pi}(\tilde{\mu}, \tilde{\nu}) \right\}$$

follows from the explicit form of Q_{opt} , described in Lemma 5.4. \square

6. DECOMPOSITION OF WASSERSTEIN-LIKE DISTANCES

In this section we discuss an applications of the decomposition theorem (Theorem 5.5) to the theory of Wasserstein-like distances.

Let $d : X \times X \rightarrow \mathbb{R}$ be a given distance function, $\text{Dom}_Q \subseteq \mathcal{P}(X)$ be an ergodic decomposable simplex. Recall, that by Wasserstein-like distance we mean the function $W_p^R : \text{Dom}_Q \times \text{Dom}_Q \rightarrow [0, \infty]$, defined by the formula (3.7).

For any $[0, +\infty]$ -valued distance function \bar{d} on the set $\partial_e \text{Dom}_Q$ of extreme points of Dom_Q define an extension of this function to the entire Dom_Q :

$$(6.1) \quad \hat{d}_p(\mu, \nu) := \inf \left\{ \left(\int \bar{d}^p(\xi, \eta)d\pi : \pi \in \Pi(\tilde{\mu}, \tilde{\nu}) \right)^{\frac{1}{p}}, \right.$$

where $\tilde{\mu}, \tilde{\nu} \in \mathcal{P}(\partial_e \text{Dom}_Q)$, $\Pi(\tilde{\mu}, \tilde{\nu}) := \{\pi \in \mathcal{P}(\partial_e \text{Dom}_Q \times \partial_e \text{Dom}_Q) : \text{Pr}_1(\pi) = \tilde{\mu}, \text{Pr}_2(\pi) = \tilde{\nu}\}$, $\text{bar}(\tilde{\mu}) = \mu$, $\text{bar}(\tilde{\nu}) = \nu$.

We are able to formulate the following statement.

Theorem 6.1 (Decomposition of a restricted Wasserstein distance). Let (X, d) be a metric Polish space equipped with Borel σ -algebra \mathcal{A} , $\text{Dom}_Q \subseteq \mathcal{P}(X)$ be a closed ergodic decomposable simplex with the associated Markov kernel Q , such that $\{Q^x\} = \partial_e \text{Dom}_Q$, \mathcal{A}^0 be a corresponding H -sufficient σ -subalgebra, $R = (\Omega, \text{Dom}_Q, \text{Dom}_Q)$ be a *geometric, ergodic decomposable*, and *coherent* linear restriction w.r.t. M , where $M \subseteq \mathcal{P}(X \times Y)$ is the corresponding ergodic decomposable simplex, Q_M and $(\mathcal{A} \otimes \mathcal{B})^0 \subseteq \mathcal{A} \otimes \mathcal{B}$ are the associated Markov kernel and H -sufficient σ -algebra.

Then for $\bar{d} := W_p^R|_{\partial_e \text{Dom}_Q}$ (the restriction of W_p^R to the set $\partial_e \text{Dom}_Q$) it is true that

$$W_p^R(\mu, \nu) = \hat{d}_p(\mu, \nu), \quad \forall \mu, \nu \in \text{Dom}_Q,$$

where \hat{d}_p is defined by the formula (6.1). Moreover, W_p^R is an $[0, +\infty]$ -valued distance function on Dom_Q .

Proof. Since $R = (\Omega, \text{Dom}_Q, \text{Dom}_Q)$ is a geometric restriction, it follows that it is weakly regular. It is known, that d^p is bounded below and is a lower semi-continuous function on $X \times X$. Hence $\pi \rightarrow \int d^p d\pi$ is a weakly lower semi-continuous functional on $\mathcal{P}(X \times X)$ (w.r.t. topology of weak convergence). Since the hypothesis of Theorem 5.5 is satisfied, one can apply it to obtain

$$\inf \left\{ \int d^p d\pi : \pi \in \Pi_R(\mu, \nu) \right\} = \inf \left\{ \int \inf \left\{ \int d^p d\pi : \pi \in \Pi_R(\xi(x), \eta(y)) \right\} d\tilde{\pi} : \tilde{\pi} \in \tilde{\Pi}(\tilde{\mu}, \tilde{\nu}) \right\}.$$

Here $\tilde{\Pi}(\tilde{\mu}, \tilde{\nu})$ is as in Definition 5.1. Since

$$(6.2) \quad (x, y) \rightarrow \left\{ \int d^p d\pi : \pi \in \Pi_R(\xi(x), \eta(y)) \right\}$$

is measurable w.r.t. $\mathcal{A}^0 \otimes \mathcal{B}^0$, and $\mathcal{A}^0 \otimes \mathcal{B}^0 \subseteq (\mathcal{A} \otimes \mathcal{B})^0$, it is true, that one can replace $\tilde{\Pi}(\tilde{\mu}, \tilde{\nu})$ with $\Pi(\tilde{\mu}, \tilde{\nu})$ without any change of the infimum:

$$\inf \left\{ \int d^p d\pi : \pi \in \Pi_R(\mu, \nu) \right\} = \inf \left\{ \int \inf \left\{ \int d^p d\pi : \pi \in \Pi_R(\xi(x), \eta(y)) \right\} d\tilde{\pi} : \tilde{\pi} \in \Pi(\tilde{\mu}, \tilde{\nu}) \right\}.$$

By the definitions of the distances \hat{d}_p , W_p^R and due to the established fact about equivalence of σ -algebras on the space of measures (Proposition 2.19), the obtained equality is equivalent to the equality

$$W_p^R = \hat{d}_p.$$

It follows from Proposition 3.11 that $W_p^R(\mu, \nu)$ is actually $[0, +\infty]$ -valued distance function on Dom_Q . \square

It can be noted, that the example of decomposition described in the introduction section appears to be a particular case of the just proved statement.

ACKNOWLEDGMENT

The paper was written during an internship of the author in Scuola Normale Superiore di Pisa. The author is grateful to SNS for its hospitality, Dr. Luigi Ambrosio for interesting and fruitful mathematical conversations, and Higher School of Economics for a sponsorship of this visit.

The author wish to thank V. I. Bogachev, A. M. Vershik, and A. V. Kolesnikov for important remarks and for the help with preparation of this paper.

REFERENCES

- [1] Aliprantis C. D., Burkinshaw O., *Positive operators*. Pure and Applied. Mathematics, vol. 119, Academic Press, New York, 1985.
- [2] Ambrosio L., Gigli N., *A user's guide to optimal transport*. Modeling and Optimization of Flows on Networks Lecture Notes in Mathematics pp 1-155, 2013.
- [3] Batty J. K., *Simplexes of states of C^* -algebras*. Volume 4, Issue 1, Summer 1980. C. J. Operator Theory pp. 3-23.
- [4] Beiglböck M., Griessler C., *An optimality principle with applications in optimal transport*. arXiv:1404.7054 [math.PR], 2014.
- [5] Bogachev V. I., *Measure Theory, Vol I. and II.* Springer-Verlag, Heidelberg, 2007.
- [6] Bogachev V. I., Kolesnikov A. V., *The Monge-Kantorovich problem: achievements, connections, and perspectives*, Russian Mathematical Surveys, 67(5):785, 2012.
- [7] Chodosh O., *Optimal Transport and Ricci Curvature: Wasserstein Space Over the Interval*. arXiv:1105.2883 [math.MG], 2011.
- [8] Colombo M., Di Marino S., *Equality between Monge and Kantorovich multimarginal problems with Coulomb cost*, Annali di Matematica Pura ed Applicata, April 2015, Volume 194, Issue 2, pp 307-320.
- [9] Cortez M.I., Rivera-Letelier J., *Choquet simplices as spaces of invariant probability measures on post-critical sets*. Ann. de l'Inst. H. Poincaré - Non Linear Analysis, Volume 27, Issue 1, 2010, Pages 95-115.
- [10] Downarowicz T., *The Choquet simplex of invariant measures for minimal flows*. Israel J. Math. 74 (1991), no. 2-3, 241-256.
- [11] Dynkin E. B., *Sufficient Statistics and Extreme Points*. Ann. Probab. 6 (1978), no. 5, 705-730. doi:10.1214/aop/1176995424, Reader in Mathematics, University of Essex, 2002.
- [12] Einsiedler M., Ward T., *Ergodic Theory: with a view towards Number Theory*. Springer Verlag 2011.
- [13] Fremlin D. H., *Measure Theory*, University of Essex, 2002.
- [14] Gaudard M., Hadwin D., *Sigma-Algebras on Spaces of Probability Measures*, Scandinavian Journal of Statistics Vol. 16, No. 2 (1989), pp. 169-175.
- [15] Himmelberg C., *Measurable relations*. Fundamenta Mathematicae (1975), Volume: 87, Issue: 1, page 53-72.
- [16] Kolesnikov A. V., Zaev D., *Optimal transportation of processes with infinite Kantorovich distance. Independence and symmetry*. arXiv:1303.7255 [math.FA], 2013.
- [17] Kerstan J., Wakolbinger A., *Ergodic decomposition of probability laws*. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete, 1981, Volume 56, Issue 3, pp 399-414.
- [18] Moameni A., *Invariance properties of the Monge-Kantorovich mass transport problem*. arXiv:1311.7051 [math.AP], 2013.
- [19] Molchanov I., *Random Closed Sets*. Space, Structure and Randomness, Lecture Notes in Statistics Volume 183, 2005, pp 135-149.
- [20] Parthasarathy K. R., *Probability Measures on Metric Spaces*. AMS Chelsea Publishing, 1967.
- [21] Phelps R., *Lectures on Choquet's Theorem*. Springer-Verlag Berlin Heidelberg, doi 10.1007/b76887, 2001.
- [22] Rieder U., *Measurable selection theorems for optimization problems*. Manuscripta mathematica, 1978, Volume 24, Issue 1, pp 115-131.
- [23] Vershik A.M., *Kantorovich Metric: Initial History and Little-Known Applications*, Journ. of Math. Sc., March 2006, Volume 133, Issue 4, pp 1410-1417
- [24] Vershik A.M., Zatitsky P.B., Petrov F.V., *Virtual continuity of measurable functions and its applications*, Russian Mathematical Surveys(2014), 69(6):1031
- [25] Vershik A.M., *Equipped Graded Graphs, Projective Limits of Simplices, and Their Boundaries*, Journ. of Math. Sc., Sep. 2015, Volume 209, Issue 6, pp 860-873
- [26] Villani C., *Optimal transport, Old and new*. Grundlehren der Mathematischen Wissenschaften, Springer-Verlag, Berlin, 2009.

- [27] Zaev D., *On the Monge-Kantorovich problem with additional linear constraints.*, arXiv:1404.4962 [math.FA], 2014.

FACULTY OF MATHEMATICS, HIGHER SCHOOL OF ECONOMICS, MOSCOW
E-mail address: `zaev.da@gmail.com`